
Archival Featured Article

Socially desirable and non-purposeful responding on the Neuroticism Extraversion Openness Personality Inventory-Revised

Larry C. Bernard ¹, R. Patricia Walsh ¹

¹Department of Psychology, Loyola Marymount University, Los Angeles, California, USA

Correspondance: Larry C. Bernard (lbernard@lmu.edu)

ABSTRACT - This study explores socially desirable and non-purposeful responding on the Neuroticism Extraversion Openness Personality Inventory - Revised (NEO PI-R; Costa & McCrae, 1992), an objective measure of the “Big Five” personality traits: Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness. The NEO PI-R's three validity-check items were examined in relation to the traditional, lengthier validity scale of another objective personality measure, the Personality Research Form (PRF; Jackson, 1999). In a sample of 237 university students, none of the NEO PI-R's validity check items were significantly correlated with the PRF Desirability. In addition, the PRF Infrequency scale (non-purposeful responding) was significantly correlated with one and the Desirability scale with four NEO PI-R scales, and PRF Desirability and Infrequency scores jointly predicted Conscientiousness scores. The implications are discussed and interestingly, may provide support for the Conscientiousness scale construct, because four of five NEO PI-R scales may be obliquely related by a higher-order factor of social desirability.

Keywords:

NEO Personality Inventory-Revised; Big Five; Social desirability; Response validity; Conscientiousness; Personality assessment; Individual differences

Introduction

There is a long tradition of concern about how respondent distortion might affect psychological traits inferred from questionnaire scores (e.g., Ben-Porath & Waller, 1992; Dunette

ISSN:
1541-745X (Print)
2169-3951 (Online)

Editor's note: This article is a reprint of an original work published in 2004 in *Counseling and Clinical Psychology Journal*, Vol. 1(1), pp. 4-16.

©2004/2018 IDR Ltd Co. This is an open access article under the Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>).



et al., 1962). This concern becomes especially relevant in clinical assessment. As a result, many questionnaires incorporate scales designed to detect respondents' distortion of responses: Eysenck Personality Questionnaire (EPQ; Eysenck & Eysenck, 1975); Minnesota Multiphasic Personality Inventory-2 (MMPI-2; Butcher et al., 1989); Personality Assessment Inventory (PAI; Morey, 1991); Personality Research Form (PRF; Jackson, 1999); and Sixteen Personality Factor Questionnaire (16PF; Cattell et al., 1982).

However, the Neuroticism Extraversion Openness Personality Inventory - Revised (NEO PI-R; Costa & McCrae, 1992), based on the "Big Five" factor model of personality (Costa & McCrae, 1988; Digman, 1990), omits "elaborate" validity scales. Some reviewers have been favorably inclined toward the NEO PI-R (e.g., Dollive, 1987; Leong & Dollinger, 1990) and one reviewer even found its lack of validity scales... "a refreshing break with the past" (Hogan, 1989, p. 547). The NEO PI-R appears to have achieved a high degree of use. In a search of the PsychINFO database from 1993 to 2002, we found the NEO PI-R used more than most other similar measures of "normal" personality traits combined (including the EPQ, PAI, 16PF, and PRF).

The NEO PI-R's authors, Costa and McCrae (1992), while acknowledging its lack of validity scales is controversial, offer it for use in counseling and clinical settings. However, they assert "...social desirability responding is not a threat to the validity of NEO PI-R scales in most cases... The use of... (social desirability) ... scales is often counterproductive (and) is not recommended" (p. 42). Socially desirable responding is not the only potential source of distortion. The NEO PI-R also does not include a measure of non-purposeful responding, such as an infrequency scale. Costa and McCrae argue that an infrequency scale may lengthen a test and "intelligent and experienced test takers may realize that highly improbable items are 'trick questions'" which could damage "the spirit of candor and cooperation between test administrator and respondent..." (1992, p. 42).

Costa and McCrae do acknowledge not every NEO PI-R profile is valid and that some individuals may deliberately distort their responses. Yet they believe any such distortion can be handled by "proper administration (and) judicious interpretation" of the instrument (1992, p. 41). McCrae and Costa (1983) have also proposed that socially desirable responding is a trait, rather than a response style, which is part of the rationale for *not* including validity scales in the NEO PI-R. Their concern is that removing the effects of socially desirable responding instead of adding to the accuracy of prediction may reduce accuracy by removing true trait variance from scores.

Yet, Costa and McCrae are still sensitive to the potential for distortion. They suggest a few quick checks relying on visual inspection. They also include "...three simple items at the bottom of the NEO PI-R answer sheet that ask respondents to tell if they have answered honestly and accurately, responded to all the items, and marked their responses in the correct places on the answer sheet" (Costa & McCrae, 1992, p. 41). Unfortunately, there are no data provided that indicates how these items may be related to respondent distortion tendencies or how successful this approach may be.

However, a recent review of the Big Five Factor Model, including the NEO-PI-R in applied settings, concluded: (1) people can distort their responses on such self-report measures; (2) applicants score more favorably than incumbents on many personality constructs; (3) intentional distortion may have a modest effect on validity; and (4) can be "corrected" for intentional distortion (Hough, 1997). Given that, and the suggestion that the NEO PI-R may be used in clinical settings for the identification of pathology and evaluation of treatment outcome (Costa & McCrae, 1992; Costa & Widiger, 1994; McCrae & Costa, 1991), Costa and McCrae's approach to detecting invalid profiles may not be sufficient. In response, post hoc validity scales have already been developed for the NEO PI-R by others (Schinka et al., 1997; Ross et al., 1997) and the briefer NEO

FFI (Scandell, 2000). Thus, the need for traditional validity scales has reasserted itself, although in the earlier mentioned search of the PsychINFO database, we found only two studies that had used these post hoc validity scales.

We are less interested in developing post hoc validity scales for the NEO PI-R than in exploring whether the three validity-check items are valid and how socially desirable – typically conceived as content responsive impression management – and non-purposeful – typically conceived of as non-content, or random – responding might be related to the NEO PI-R scales. Because the NEO PI-R does not include such scales, there is very little information available in the literature. Skinner et al. (1976, as cited in Costa and McCrae, 1985) reported significant, but weak, correlations between the Personality Research Form E (PRF; Jackson, 1984) Infrequency scale and two NEO PI scales: Neuroticism ($r = .15, p < .01$) and Openness ($r = -.15, p < .05$). Costa & McCrae (1988) also reported the results of a factor analysis in which the PRF Desirability scale had moderate factor loadings (in parentheses) on NEO-PI Neuroticism (-.35), Extraversion (.45), and Conscientiousness (.54) scales. These relationships suggest that socially desirable and non-purposeful responding are related to scores on several of the NEO PI-R scales. We found no other recent studies that looked at the relationships between traditional scales indicating socially desirable or non-purposeful responding and the NEO PI-R scales or explored the relationships between socially desirable and non-purposeful responding, as traditionally assessed, and the NEO PI-R's three validity-check items.

We hypothesized that socially desirable responding, as measured on PRF, would be correlated with the NEO-PI-R scales. We choose the PRF validity scales because they had been used by Costa and McCrae (1985, 1988). The PRF Desirability scale represents the traditional approach in which a validity scale is developed for concurrent use with a personality assessment device (see “Materials” below for details). High scores on PRF Desirability reflect a favorable impression given in responses to personality statements, and PRF Desirability mean scores rise in response to instructions to create the best impression (Jackson, 1999).

The *NEO-PI-R Professional Manual* (Costa & McCrae, 1992) describes Neuroticism as contrasting adjustment with maladjustment; Extraversion as sociability, liking people, preferring large groups, assertiveness, activity, and talkativeness; Openness as openness to experience, general curiosity; Agreeableness as altruistic, sympathetic toward and eager to help others; and Conscientiousness as self-control, planning, organizing, purposeful, and scrupulous. We believe Neuroticism would be perceived as socially undesirable, while Extraversion, Agreeableness, and Conscientiousness would be perceived as socially desirable traits. Based on these descriptions and review of the scales' item content, we predicted PRF Desirability would be correlated negatively with Neuroticism and positively with Extraversion, Agreeableness, and Conscientiousness. Given the description of Openness as curious and open to experience, we did not make a prediction for the Openness domain.

High scores on the PRF Infrequency scale reflect pseudo-random responding due to “carelessness, poor comprehension, passive non-compliance, confusion, or gross deviation” (Jackson, 1999, p. 7). Poor comprehension, confusion, or gross deviation are unlikely in a college population, therefore high Infrequency scores would likely indicate carelessness or non-purposeful responding. Given a reasonable level of motivation and interest in a sample or university students, and that the PRF Infrequency scale reflects non-content responding, we did not expect it to be more than weakly correlated with four of the NEO PI-R domain scales. However, because the Conscientiousness scale is intended to tap purposeful and scrupulous behavior, we hypothesized that the PRF Infrequency scale would be negatively correlated with it. We expected participants with high Conscientiousness scores to take care to respond to item content and

therefore to have lower scores on the PRF Infrequency scale. Given the predicted relationships between Conscientiousness and the PRF Desirability and Infrequency scales, we also hypothesized that both PRF scales would jointly predict Conscientiousness scores.

Our final, central hypothesis for clinical purposes involved the NEO PI-R's three "yes/no" validity-check items: 1) "I answered honestly and accurately;" 2) "I responded to all the statements;" and 3) "I answered in correct areas." Costa and McCrae (1992) suggest these items will help detect distortion. On the face of it, endorsing them, particularly the first, should indicate a socially responsible, hence socially desirable, approach to the task. Therefore, we expected them to correlate with the PRF Desirability scale. We made no prediction regarding the PRF Infrequency scale because it did not seem logical that a person who did not respond purposefully (i.e., did not respond to the content of NEO PI-R items) would suddenly respond purposefully to the three validity-check items.

Method

Participants

Data were collected as part of another study of university subject pool participants (Bernard & Walsh, 2002). A total of 237 people, 79 men and 157 women, participated, with one person not reporting sex (and consequently dropped). All participants were treated in accord with American Psychological Association standards for ethical treatment and received partial course credit for participation. All data were collected anonymously. The average age of participants was 19.17 years ($SD = 3.17$). There were 146 (61.6%) Euro American participants, 45 (19.1%) Latino/a participants, 12 (5%) African American participants, 20 (8.4%) Asian/Pacific participants, and 14 (5.9%) who marked "other."

There are significant mean sex differences in both the PRF and NEO PI-R normative samples for at least some scales. For example, female samples tend to score higher than male samples on NEO PI-R Neuroticism and Agreeableness (Costa & McCrae, 1992). In addition, separate profile forms based on mean sex differences are available for the PRF and NEO PI-R. Therefore, we expected and found mean sex differences in the present sample (e.g., the male sample's mean PRF Infrequency score was significantly higher than the female sample's mean, $p < .05$) and therefore separated the male and female samples for the present analysis.

Materials

The *Neuroticism, Extraversion, Openness Personality Inventory – Revised*, Form S (NEO PI-R, Costa & McCrae, 1992) is a 240-item measure of the "Big Five" personality domains: Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness. The scales were constructed from a rationally guided factor analytic process. Each domain scale is comprised of six facets, subscales, that tap more circumscribed traits related to the domain. The NEO PI-R Manual (Costa & McCrae, 1992) reports good internal consistency for the domain scales ($\alpha = .86$ to $.92$) and good retest reliability, ranging from $r = .75$ to $.92$ for earlier versions of the Neuroticism, Extraversion, and Openness scales, and from $r = .75$ to $.83$ for shorter versions of all five domain scales. Stability coefficients from a seven-year longitudinal study of peer ratings ranged from $.63$ to $.81$ for the five domain scales. The Manual also reports several studies attesting

to the NEO PI-R's convergent, divergent, and criterion validity. In particular, the authors have made extensive use of observer ratings as objective validity criteria.

The Personality Research Form, Form E (PRF, Jackson, 1999) was derived from Henry Murray's (1938) concept of normal personality but unlike the NEO PI-R, it was not developed factor analytically. Instead, it consists of 22 scales representing "...characteristics judged to be most important or most relevant to a wide variety of areas of human functioning..." (Jackson, 1999, p. 2). In contrast to the NEO PI-R, the PRF includes two 20-item validity scales: Desirability and Infrequency. The Desirability scale was developed to elicit consistent tendencies to respond desirably or undesirably while avoiding psychopathological item content, including heterogeneous item content and containing content similar to other PRF scales. The scale has good internal consistency as indicated by a Kuder-Richardson 20 reliability estimate of .83. One- and two-week test-retest reliabilities of .81–.86 have been reported in several studies for Desirability (see Jackson, 1999). The Desirability items were included on the basis of a set of systematic steps which included assessing styles of responding rather than content. The Manual reports that two alternative forms load a common factor and are "distinguishable from content scales" (Jackson, 1999, p. 37). It also reports that instructions to fake under various instructional sets produces expected changes in the mean Desirability score.

The PRF "Infrequency scale was designed to identify careless, non-purposeful and invalid responding" (Jackson, 1999, p. 37). Items were written to identify behavior that was either extremely common or uncommon in the general population, and the author claims "it was possible to identify a class of content which, while very deviant statistically, is neither bizarre nor particularly undesirable" (Jackson, 1999, p. 37). A purposeful keyed response to an Infrequency item would be extremely rare. The means of the PRF college normative sample on Form E for the male and female sample, respectively were 0.48 ($SD = .72$) and 0.37 ($SD = .78$), indicating a very low endorsement rate. Despite the skewed score distribution that this implies, early versions of the 65-item pool from which items were ultimately selected had a Kuder-Richardson 20 reliability estimate of .52. One- and two-week test-retest reliabilities of .46–.77 have been reported in several studies for Infrequency (see Jackson, 1999). The reliability estimate is typically lower than for other PRF scales and for Desirability and taken by the author as "supporting the hypothesized presence of random response tendencies" (see Jackson, 1999, p. 38). The Manual reports that the scale "is highly sensitive to certain types of invalid responding..." (Jackson, 1999, p. 38).

Procedure

The present study was titled "Characteristics of Subject Pool Participants." Participants were run during the 4th, 5th, 14th, and 15th weeks of a 16-week semester, to reduce temporal sampling effects (Bernard, 2000).

Results

Table 1 depicts the means and standard deviations of males and females in the present sample and in the PRF College Nonnative Group. Compared to the College Normative Group, male participants in the present sample had significantly higher mean Desirability and Infrequency

scores, and female participants in the present sample had significantly higher Infrequency but not Desirability scores. There was a significant correlation between the PRF Desirability and infrequency scales in this sample for males of $r = -.41$ ($p < .001$) and for females or $r = -.23$ ($p = .004$). In a logically consistent manner, higher scores on Desirability were associated with less non-purposeful responding.

Table 1: Means and standard deviations of males and females in the present sample and the College Normative Sample on Personality Research Form Desirability and Infrequency scales

	Present Sample		PRF Form E College Normative Sample	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Infrequency				
Males**	2.04	2.68	0.48	0.72
Females**	0.83	1.37	0.37	0.78
Desirability				
Males*	10.08	2.83	10.78	3.00
Females	10.86	2.78	10.97	2.53

Note: For the present sample, N males = 79 and N females = 157; for the PRF Form E college normative sample, N males = 1,350, N females = 1,415. The significance of differences in means is: * $p = .03$ ** $p = .0001$.

The central hypothesis for clinical purposes involved the NEO PI-R's three "yes/no" validity-check items: 1) "I answered honestly and accurately;" 2) "I responded to all the statements;" and 3) "I answered in correct areas." We computed the zero order correlations between the PRF Desirability scale, and the three NEO PI-R validity-check items as depicted in Table 2. None of the correlations were significant ($p > .05$).

Table 2: Zero order correlations between the Personality Research Form Desirability scale and the NEO Personality Inventory-Revised's three validity-check items

	Answered honestly and accurately	Responded to all the statements	Answered in correct areas
Male sample	.08	-.04	— ¹
Female sample	.19	.06	.03

Note: None of the correlation coefficients were significant ($p > .05$). ¹Could not be computed due to invariance of responses.

The results of zero order correlations between the PRF validity scales and NEO PI-R scales are presented in Table 3. Each of the PRF validity scales was considered to have a separate hypothetical relationship to the five NEO PI-R scales. A Bonferroni correction was applied to reduce the probability of Type I error inflation, and corrected $\alpha = .01$ [$.05 \div 5$] was used as the criterion for significance.

Table 3: Zero order correlations between Personality Research Form Desirability and Infrequency scales and NEO Personality Inventory-Revised domain scales

NEO Personality Inventory- Revised domain scales	Personality Research Form			
	Infrequency		Desirability	
	Males	Females	Males	Females
Neuroticism	.19	-.04	-.46**	-.52**
Extraversion	-.31*	.04	.45**	.43**
Openness	-.31*	-.01	.14	.17
Agreeableness	-.20	-.10	.38*	.39**
Conscientiousness	-.41**	-.26*	.43**	.40**

Note: Males, $N = 79$; Females $N = 157$. * $p < .01$ ** $p < .001$

Table 4: Multiple regressions of Personality Research Form Desirability and Infrequency scales on the Neuroticism Extraversion Openness Personality Inventory-Revised Conscientiousness scale

	B	$SE\ B$	β
Male sample			
Desirability	2.309	0.884	.289*
Infrequency	2.346	0.922	-.282*
Female sample			
Desirability	2.483	0.558	.336**
Infrequency	2.808	1.168	-.181

Note: Males $R^2 = .481$ ($p = .0001$), Females $R^2 = .481$ ($p = .0001$). * $p < .01$ ** $p < .001$

The PRF Desirability scale had a significant, moderate correlation with four of the five NEO PI-R domain scales for both males and females: Neuroticism, Extraversion, Agreeableness, and Conscientiousness. For males, the PRF Infrequency scale had a significant, moderate correlation with three of the five NEO PI-R domain scales: Extraversion, Openness, and Conscientiousness. For females, the PRF Infrequency scale had only one significant, weak correlation with NEO PI-R Conscientiousness.

The results of the multiple regressions of the PRF Desirability and Infrequency scales on NEO PI-R Conscientiousness are depicted in Table 4 for males and females, respectively. Both PRF Desirability and Infrequency emerged as significant independent predictors of Conscientiousness for both males and females.

Discussion

We explored the relationship between the NEO PI-R's three validity-check items and the PRF's more traditional Desirability scale. Only the first validity check item – "I answered honestly and accurately" – has strong face-validity in relation to social desirable responding, however, neither it, nor either of the other two items, had significant relationships with PRF Desirability. Therefore, the three NEO PI-R validity-check items may not be sufficiently sensitive to detect the effects of socially desirable responding.

However, socially desirable and non-purposeful responding, as measured by the PRF's Desirability and Infrequency scales, do appear to be related to at least some of the NEO PI-R domain scales. Desirability was significantly correlated with four of the NEO PI-R's five domain scales. The pattern of correlations – negative with Neuroticism and positive with Extraversion, Agreeableness, and Conscientiousness – were the same for male and female participants. This is not the first time such relationships have been found. McCrae and Costa (1983; see also McCrae, 1986) reported similar relationships (particularly with regard to Neuroticism, but with the other domains as well) when they partialled out scores on a measure of social desirable responding from the correlations between self-reporting and ratings by others on the NEO. They argued that partial correlations between the self-rating and the more objective criterion rating by others should have increased if socially desirable responding distorted the self-rated responses. But the partial correlations decreased slightly, and McCrae and Costa used that observation to argue that socially desirable responding and Neuroticism are overlapping traits. The same argument could be made for Conscientiousness.

However, as reasonable as that argument is, in the present study, in addition to Neuroticism and Conscientiousness, socially desirable responding was correlated with two other NEO PI-R scales. If socially desirable responding were consistently found to be correlated with four of five NEO PI-R domain scales, it would raise interesting questions, if not about the validity of the NEO PI-R scales, then about the relationship of the constructs in the Big Five Model. If we accept that socially desirable responding is a substantive trait and not a style, and if it is potentially correlated with four of the Big Five factors on the NEO PI-R, that would imply the existence of a higher-order factor in NEO PI-R scales and possibly the Big Five Model of personality. This higher-order factor would be the trait of social desirability that would partially influence the variance in four of the Big Five factors as measured by the NEO PI-R. That would suggest an oblique relationship among the NEO PI-R scales, while the theory that defines the Big Five factor model and the factor-analytic methods used to develop the NEO PI-R and other Big Five measures posit orthogonal traits. This is not necessarily a problem for the NEO PI-R or the Big Five model, for it does support Costa and McCrae's argument that socially desirable responding is a personality trait in its own right. When conceptualized as a trait, and not a nuisance in assessment, social desirability may have intriguing implications for understanding personality dynamics in clinical and counseling. Perhaps, then, the role of socially desirable responding in the Big Five model needs to be addressed more.

With regard to other results, PRF Infrequency was negatively correlated with Conscientiousness for male and female participants. Finally, we explored the extent to which PRF

Desirability and Infrequency scales could serve as joint predictors of NEO PI-R Conscientiousness. Desirability and Infrequency both made significant, independent contributions to the prediction of Conscientiousness scores for male and female participants. That is not surprising, but of interest is that adjusted R^2 indicated that about 21% and 16% of the variance in male and female participants' Conscientiousness scores, respectively, may be accounted for by their Desirability and Infrequency scores.

Costa and McCrae made a deliberate decision not to include validity scales in the NEO PI-R (Costa & McCrae, 1992). First, as mentioned, they suggested that socially desirable responding might be a trait, the effects of which, if removed from the NEO PI-R scales, could reduce the correlations between them and other criteria (McCrae and Costa, 1983). How much of an effect would adjusting for socially desirable responding have on the NEO PI-R? The coefficients of determination (r^2) in the present sample indicate that PRF Disability and the five NEO PI-R domain scales share between 14–27% of their variance in this sample, suggesting that up to about a quarter of the variance in four of the NEO PI-R domain scales would be affected. That amount of variance could have an impact on interpretation of NEO PI-R scores. So, the decision to include, or not, an adjustment for socially desirable responding needs to be made in practice with careful deliberation. While counselors and clinicians should probably not rely on the NEO PI-R's three validity-check items, they should consider the role socially desirable responding may play in scores on the five domain scores as well as its possible role as a personality trait itself.

Second, Costa and McCrae believe that an infrequency scale may lengthen a test with little or no advantage and compromise the respondent's desire to cooperate. However, experimental studies have indicated that manipulation of participants' responding – either non-purposeful or positive or negative impression management – can affect scores on the NEO PI-R (Schinka et al., 1997; Ross et al., 1997) and the brief NEO-FFI (Scandell, 2000). Given that knowledge, and the results of the present study, concern about response bias on the NEO PI-R should also be given careful deliberation. This concern becomes particularly important when the NEO PI-R is used in clinical and applied settings. On the other hand, by actively engaging the study of response distortion on the NEO PI-R, we have found evidence for the validity of at least one scale, Conscientiousness, supporting the utility of this line of research.

That said, we must caution against over generalization from this study, because this sample may not be equivalent to other college samples, nor to a clinical sample (the NEO PI-R Manual does not offer a clinical sample with which to compare performance). Compared to the PRF Form E College Normative Sample, this sample's male and female participants did have significantly higher mean Infrequency scores. Although significant, the female participants' mean, at about 1 ($SD = 1.4$), does not differ from the Normative Sample by even one raw score unit, but the male participants' mean at about 2 ($SD = 2.7$), is a unit and a half above the College Normative Sample's mean. Although elevated, it is not above the raw score cutoff of ≥ 4 for invalid profiles recommended in the PRF manual (Jackson, 1999). While the degree of variance in Infrequency scores found in the present male sample is useful for detecting relationships, it does suggest that at least some male participants in this sample took the task less seriously than the PRF College Nonnative Sample. As for Desirability, there was not a significant difference in means between

the present sample's female participants and the PRF College Nonnative Sample's participants. There was a significant mean difference for the male participants, however. Male participants in the present sample had a significantly lower mean Desirability score of about 10 ($SD = 2.8$) compared to the PRF College Normative Sample's mean of about 10.8 ($SD = 3.0$), but again this is not even a full raw score unit difference and probably would not be meaningful in a clinical sense.

In conclusion, socially desirable and non-purposeful responding did account for at least some variance in most of the NEO PI-R domain scales in this sample. Because of that, we believe the concerns of two reviewers (Hess, 1992; Widiger, 1992) about the NEO PI-R's lack of validity scales and its assumption of a cooperative, honest subject, which led them to question the NEO PI-R's use in forensic, personnel, and clinical settings, should be accorded more investigation. Additional research on the effects of socially desirable and non-purposeful responding on the NEO PI-R could shed more light on the questions raised here.

References

- Ben-Porath, Y. S., & Waller, N. G. (1992). "Normal" personality inventories in clinical assessment: General requirements and the potential for using the NEO Personality Inventory. *Psychological Assessment*, 4(1), 14–19. <https://doi.org/10.1037/1040-3590.4.1.14>
- Bernard L. (2000). Variations in subject pool as a function of earlier or later participation. *Psychological reports*, 86(2), 659–668. <https://doi.org/10.2466/pro.2000.86.2.659>
- Bernard, L. G., & Walsh, R. P. (2002). Variations in a university subject pool as a function of earlier or later participation and self-report: A replication and extension. *Psychological Reports*, 91(2), 553–570. <https://doi.org/10.2466/PRO.91.6.553-570>
- Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *Minnesota Multiphasic Personality Inventory-2 (MMPI-2). Manual for Administration and Scoring*. University of Minnesota Press.
- Cattell, R. B., Eber, H.W., & Tatsuoka, M. M. (1982) *Handbook for the Sixteen Personality Factor Questionnaire*. Institute for Personality and Ability Testing.
- Costa, P., & McCrae, R. (1985). *The NEO Personality Inventory Manual*. Psychological Assessment Resources.
- Costa, P. T., Jr., & McCrae, R. R. (1988). From catalog to classification: Murray's needs and the five-factor model. *Journal of Personality and Social Psychology*, 55(2), 258–265. <https://doi.org/10.1037/0022-3514.55.2.258>
- Costa, P. T., Jr., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Psychological Assessment Resources.
- Costa, P. T., Jr., & Widiger, T. A. (Eds.). (1994). *Personality disorders and the five-factor model of personality*. American Psychological Association. <https://doi.org/10.1037/10140-000>
- Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology*, 41, 417–440. <https://doi.org/10.1146/annurev.ps.41.020190.002221>

- Dollive, R. H. (1987). A review of the NEO Personality Inventory. *Journal of Counseling & Development*, 66(2), 107–108. <https://doi.org/10.1002/j.1556-6676.1987.tb00813.x>
- Dunnette, M. D., McCartney, J., Carlson, H. C., & Kirchner, W. K. (1962). A study of faking behavior on a forced-choice self-description checklist. *Personnel Psychology*, 15(2), 13–24. <https://doi.org/10.1111/j.1744-6570.1962.tb01843.x>
- Eysenck, H. J., & Eysenck, S. B. G. (1975). *Manual of the Eysenck Personality Questionnaire (junior and adult)*. Hodder and Stoughton.
- Hess, A. K. (1992). Review of the NEO Personality Inventory. In J. C. Conoley & J. J. Kramer (Eds.). *The eleventh mental measurements yearbook* (pp. 603–605). Burns Institute of Mental Measurements.
- Hogan, R. (1989). Review of the NEO Personality Inventory. In J. C. Conoley & J. J. Kramer (Eds.). *The tenth mental measurements yearbook* (pp. 546–547). Buros Institute of Mental Measurement.
- Hough, L. M. (1998). The millennium for personality psychology: New horizons or good ole daze. *Applied Psychology: An International Review*, 47(2), 233–261. <https://doi.org/10.1111/j.1464-0597.1998.tb00023.x>
- Jackson, D. N. (1984). *Personality Research Form manual* (3rd ed.) Research Psychologists Press.
- Jackson, D. N. (1999). *Personality Research Form manual* (3rd ed. – updated) Research Psychologists Press.
- Leong, F. T. L., & Dollinger, S. J. (1990). NEO Personality Inventory. In D. J. Keyser & R. C. Sweetland (Eds.), *Test Critiques* (Vol. VIII, pp. 527–539). Austin, TX: PRO-ED.
- McCrae, R. R. (1986). Well-being scales do not measure social desirability. *Journal of Gerontology*, 41(3), 390–392. <https://doi.org/10.1093/geronj/41.3.390>
- McCrae, R. R., & Costa, P. T. (1991). The NEO Personality Inventory: Using the Five-Factor Model in counseling. *Journal of Counseling & Development*, 69(4), 367–372. <https://doi.org/10.1002/j.1556-6676.1991.tb01524.x>
- McCrae, R. R., & Costa, P. T. (1983). Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology*, 51(6), 882–888. <https://doi.org/10.1037/0022-006X.51.6.882>
- Morey, L. C. (1991). *The Personality Assessment Inventory: Professional manual*. Psychological Assessment Resources.
- Murray, H.A. (1938). *Explorations in personality*. Oxford University Press.
- Ross, S. R., Bailey, S. E., & Millis, S. R. (1997). Positive self-presentation effects and the detection of defensiveness on the NEO PI–R. *Assessment*, 4(4), 395–408. <https://doi.org/10.1177/107319119700400410>
- Scandell, D. J. (2000). Development and initial validation of validity scales for the NEO-Five Factor Inventory. *Personality and Individual Differences*, 29(6), 1153–1162. [https://doi.org/10.1016/S0191-8869\(99\)00262-7](https://doi.org/10.1016/S0191-8869(99)00262-7)

- Schinka, J. A., Kinder, B. N., & Kremer, T. (1997). Research validity scales for the NEO-PI-R: development and initial validation. *Journal of Personality Assessment*, 68(1), 127–138. https://doi.org/10.1207/s15327752jpa6801_10
- Skinner, H. A., Jackson, D. N., & Rampton, G. M. (1976). The Personality Research Form in a Canadian context: Does language make a difference? *Canadian Journal of Behavioural Science / Revue canadienne des sciences du comportement*, 8(2), 156–168. <https://doi.org/10.1037/h0081944>
- Widigcr, T. A. (1992). Review of the NEO Personality inventory. In J. C. Conoley & J. J Kramer (Eds.), *The eleventh mental measurements yearbook* (pp. 605–606). Burns Institute of Mental Measurements.